

SEQUENCE LISTING

(1) GENERAL INFORMATION

- (i) APPLICANT: Jung, Rudolf
Beach, Larry R.
Dress, Virginia M.
Rao, A. Gururaj
Ranch, Jerome P.
Ertl, David S.
Higgins, Regina K.
- (ii) TITLE OF THE INVENTION: Alteration of Amino Acid Compositions
in Seeds
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Pioneer Hi-Bred International, Inc.
(B) STREET: 7100 NW 62nd Avenue, P.O. Box 1000
(C) CITY: Johnston
(D) STATE: IA
(E) COUNTRY: USA
(F) ZIP: 50131
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Diskette
(B) COMPUTER: IBM Compatible
(C) OPERATING SYSTEM: DOS
(D) SOFTWARE: FastSEQ for Windows Version 2.0
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER:
(B) FILING DATE:
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Michel, Marianne H
(B) REGISTRATION NUMBER: 35,286
(C) REFERENCE/DOCKET NUMBER: 0815
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: 515-334-4467
(B) TELEFAX: 515-334-6883
(C) TELEX:

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 3363 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

TCGACCTCGA	GGGGGGGCCC	GGTACCAGC	TTTTGTTC	TTAGTGAGG	GTAAATTGCG	60
CGCTTGGCGT	AATCATGGTC	ATAGCTGTTT	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	120
CCACACACAA	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGGTGCCTA	ATGAGTGAGC	180
TAACCTACAT	TAATTGCGTT	CGGCTCACTG	CCCGCTTTCC	AGTCGGGAAA	CCTGTGCTGC	240
CAGCTGCATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCGTAT	TGGGCGCTCT	300
TCGCTTCTCT	CGCTCACTGA	CTCGCTGCGC	TCGGTCTGTC	GGCTGCGGCG	AGCGGTATCA	360
GCTCACTCAA	AGGCGGTAA	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	420
ATGTGAGCAA	AAGGCCCTAA	AAAGGCCAGG	AACCGTAAAA	AGGCCGCGTT	GCTGGCGTTT	480
TTCCATAGCG	TCCGCCCCCC	TGACGAGCAT	CGGTTTCCCC	GACGCTCAAG	TCAGAGGTGG	540
CGAAACCCGA	CAGGACTATA	AAGATACCAG	TACCTGTCCG	GCTGCGCCTT	TTGCGGAAGC	600
TCCTCTGTTC	CGACCTGCTG	GCTTACCGGA	TATCTCAGTT	CGGTGTAGGT	CGTTCGCTTC	660
GTGGCGCTTT	CTCATAGCTC	ACGCTGTAGG	CAGCCCCGAC	CACTGGCAGC	AGGCACCTGGT	720
AAGCTGGGCT	GTGTGCACGA	ACCCCCGCTT	GACTTATCGC	CACTGGCAGC	ATCCGGTAAAC	780
TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GCTGTTCCCC	CGTTCGCTTC	ATCCGGTAAAC	840
AACAGGATTA	CGCAGAGCGG	GTATGTAGGC	GATGTTAGGC	CGTTCGCTTC	ATCCGGTAAAC	900
TAACACGCGT	ACACTAGAGG	GACAGTATTT	GGTATCTGCG	CTCTGTGTAA	GCCAGTTACC	960
TTTCGGAATA	GAGTTGGTAG	CTCTTGATCC	GGCAACACAA	CCACCGGTGT	TAGCCGGTGT	1020
TTTTTTGTTT	CGAAGCAGCA	GATTACGCGC	AGAAAAAAGG	GATCTCAAGA	AGATCTCTTG	1080
ATCTTTTCTT	CGGGGTCTGA	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGCT	1140
ATGAGATTAT	CAAAAAGGAT	CTTCACCTAG	ATCCTTTTAA	ATTAAAAAAG	AAGTTTAAAG	1200
TCRAATCTAA	GTATATATGA	GTAACCTTGG	CTGACAGTTT	ACCAATGCTT	AATCAGTGAAG	1260
GCACCTATCT	CAGCAGTCTG	TCTATTTCTG	TCATCCATAG	TTGCTGTGAC	CCCCGTCTGT	1320
TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGCCCCCA	GTGCTGCAAT	GATACCGCGA	1380
GACCCACGCT	CACCGGCTCC	AGATTATATCA	GCAATAAACG	AGCCAGCCCG	AAGGGCCGAG	1440
CGCAGAAAGT	CTCTCCGAAC	TTTATCCGCG	TCCATCCAGT	CTATTAAATT	TTGCCGGGAA	1500
GCTAGAGTAA	GTAGTTCCGC	AGTTAATAGT	TTGCGCAACG	TTGTTGCCAT	TGCTACAGCG	1560
ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATCTA	GCTCCGGTTC	CCAACGATCA	1620
AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	TTAGCTCCTT	CGGTCTCTCG	1680
ATCGTTGTCA	GAAGTAAGTT	GGCCGCACTG	TTATCACTCA	TGGTTATGCG	AGCACTGCAT	1740
AATTCTCTTA	CTGTCTATGCC	ATCCGTAAGA	TGCTTTTCTG	GACTGTGTGA	GTACTCAACC	1800
AAGTCATTCT	GAGAATAGTG	TATGCGGCGA	CCGAGTTGCT	TCATTGGAAA	ACGTTCTTCT	1860
GATAAATCCG	CGCCACATATG	CAGAAGCTTA	TTGAGATCCA	GTTCGGGTGT	TTTCTGGGTG	1920
GGGGCGAAAA	TCTCAAGGAT	CTTACCGCTG	TTGAGATCCA	GGAAATGTGT	AATATCTATA	1980
GCACCAAACT	GATCTTCAGC	ATCTTTTACT	TTACAGGGTT	ATTGTCTCAT	GAGCGGATAC	2040
GGAAAGGCAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	CGCGCACATT	TCCCGGAAAA	2100
CTCTTCTCTT	TTCAATATTA	TTGAAGCATT	TATGAGGTTT	CGCGTTAAAT	TTTTGTTAAA	2160
ATATTGGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTT	TGTTAAAAAT	CCCTTATAAA	2220
GTGCCACCTA	AATTGTAAGC	GTTAATATTT	TCGGCAAAAT	CCCTTATAAA	TCAAAGAAGT	2280
TCAGCTCATT	TTTTAACCAA	TAGSCCGAAA	TTTGGAACAA	GAGTCCACTA	CTTACAGAAC	2340
AGACCGAGAT	AGGGTGTAGT	GTGTTCCAG	TTTGGAACAA	GATGCGCCCA	GTACGTGAAC	2400
TGGACTCCAA	CGTCAAAAGG	CGAAAAACCG	TCTATCAGGG	CGATGCGCCA	CTAAGTGAAC	2460
CATCACCCCTA	ATCAAGTTTT	TTGGGGTCTG	GGTGCCGTAA	AGCACTAAAT	CGGAACCCCTA	2520
AAGGGAGGCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGCG	GAACGTGCGC	AGAAAGGAAG	2580
GGAAAGAAAG	GAAAGTAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
TAAACCAACC	ACCCGCGCGG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TGCCCAATTA	2700
GGCTGCGCAA	CTGTTGGGAA	GGGCGATCCG	TGCGGGCCTC	TTGCGTATTA	CGCGAGCTGG	2760
CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTGCGGTAAC	GCCAGGGTTT	TCCAGTCTAC	2820
GACGTTGTAA	AACGACGCGC	AGTGAGCGCG	CGTAATACGA	CTCACTATTG	GCGCAATTGG	2880
AGCTCCACCG	CGGTGGGCGG	CGCTCTAGAA	CTAGTGGATG	CGTCGACTAG	AGGGCCCGAC	2940
GTGCAACTTA	GGCACTAAGG	GATGTGAGCG	CAGCATCACC	GTTCGACAAA	TGACACAGAG	3000
CATCACCCAA	ATTTTCCAAA	TAGAGTTTCA	TTTCTTCGTC	GTCAGCAGCT	GCGTTGACCA	3060

TGTAAGTACCA	CATGGAAGCC	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGTTTCAT	3120
CTGAGTTGGA	CACAAGGCC	AATTGGGGA	AGCCTGTAGG	GCATTTTCCG	CTACTTTGTA	3180
GTTTACACTT	ACAGACGGCT	CGCATAACT	TCTGAGCACC	ACGGACGGG	CAAAAGTTGT	3240
AGCAGTTTCT	TCTTAGGGTG	CTCTGCGAGC	AACCTTTGCC	TTCTACTTGC	ACCTGTTCCA	3300
GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCTTT	GAGGCCCTTG	CTGGTGCCCA	3360
TGG						3363

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3365 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

TCGACCTCGA	GGGGGGGCC	GGTACCCAGC	TTTTGTTCCC	TTTAGTGAGG	GTAAATTGCG	60
CGCTTGGCGT	AATCATGTGC	ATAGCTGTTT	CCTGTGTGAA	ATTGTTATCC	GCTCACAATT	120
CCACACAACA	TACGAGCCGG	AAGCATAAAG	TGTAAAGCCT	GGGTGCTCTA	ATGAGTGAGC	180
TAATCATCAT	TAATTGCGTT	CGGCTCACTG	CCCGCTTTCC	AGTCGGGAAA	CCTGTCGTGC	240
CAGTCGATT	AATGAATCGG	CCAACGCGCG	GGGAGAGGCG	GTTTGCCTAT	TGGGCGCTCT	300
TCCGCTTCCT	CGCTCACTGA	CTCGCTGCGC	TCGGTCTGTT	GGCTGCGGCG	AGCGGTATCA	360
GCTCACTCAA	AGGCGGTAA	ACGGTTATCC	ACAGAATCAG	GGGATAACGC	AGGAAAGAAC	420
ATGTGAGCAA	AAGGCCAGCA	AAAGGCCAGG	AACCGTAAAA	AGGCCCGGTT	GCTGGCGTTT	480
TTCCATAGGC	TCCGCCCCCC	TGACGAGCAT	CACAAAAATC	GACGCTCAAG	TCAGAGGTGG	540
CGAAACCGGA	CAGGACTATA	AAGATAACAG	CGGTTTCCCC	CTGGAAGCTC	CCTCTGCGCG	600
TCTCTGTTTC	CGACCTCGCC	GCTTACCGGA	TACCTGTCCG	CCTTCTCCCC	TTCCGGGAAGC	660
GTGGCGCTTT	CTCATAGTCT	ACGCTGTAGG	TATCTCAGTT	CGGTGTAGGT	CGTTGCTCTC	720
AAGCTGGGCT	GTGTGACAGA	ACCCCCCGTT	CAGCCCCGAC	GCTGCGCCTT	ATCCGGTAAC	780
TATCGTCTTG	AGTCCAACCC	GGTAAGACAC	GACTTATCCG	CACTGGCAGC	AGCCACTGGT	840
AACAGSATT	GCAGAGCGAG	GTATGTAGGC	GGTGCTACAG	AGTTCTTGAA	GTGGTGGCCT	900
AACTACGGCT	ACACTAGAAG	GACAGTATTT	GGTATCTGCG	CTCTGCTGAA	GCCAGTTACC	960
TTCCGAAAAA	GAGTTGGTAG	CTCTTGATCC	GGCAAAACAA	CCACCGCTGG	TAGCGGTGGT	1020
TTTTTTTGTT	GCAAGCAGCA	GATTACGCGC	AGAAAAAAG	GATCTCAAGA	AGATCCTTTT	1080
ATCTTTTCTA	CGGGGTCTGA	CGCTCAGTGG	AACGAAAACT	CACGTTAAGG	GATTTTGGTC	1140
ATGAGATTAT	CAAAAAGGAT	CTTCACTTAG	ATCCTTTTAA	ATTAATAAGT	AAGTTTAAAA	1200
TCAATTATAA	GTATATATGA	GTAACCTTGG	TCGTGACAGT	ACCAATGCTT	AATCAGTGAG	1260
GCACCTATCT	CAGCGATCTG	TCTATTTGCT	TCATCCATAG	TTGCTTGACT	CCCGCTCGTG	1320
TAGATAACTA	CGATACGGGA	GGGCTTACCA	TCTGGCCCCA	GTGCTGCAAT	AGCCAGCCGG	1380
GACCCACGCT	CACCGGCTCC	AGATTATCA	GCAATAAAC	AGCCAGCCGG	AAGGGCCGAG	1440
CGCAGAAAGT	GTCTTGCAAC	TTTATCCGCC	TCCATCCAGT	CTATTAATTG	TTGCCGGGAA	1500
GCTAGAGTAA	GTAGTTCCGC	AGTTAATAGT	TTGCGCAACG	TTGTTGGCAT	TGCTACAGGC	1560
ATCGTGGTGT	CACGCTCGTC	GTTTGGTATG	GCTTCATTCA	GCTCCGGTTC	CCAACGATCA	1620
AGGCGAGTTA	CATGATCCCC	CATGTTGTGC	AAAAAAGCGG	TTAGCTCTCT	CGGTCTCTCG	1680
ATCGTTGTCA	GAAATAGTGT	TATGCGCGCA	TTTACTACTCA	TGGTTATGGC	AGCAGTCGAT	1740
AATCTCTTTA	CTGTGATGCC	ATCCGTAAGA	TGCTTTTCTG	TGACTGATGTA	GTACTCAACC	1800
GAGTCATTCT	GAGAATAGTG	TATGCGCGCA	CCGAGTTGCT	CTTGCCCGGC	GTCATACAGG	1860
AATAATACCG	CGCCACATAG	CAGAACTTTA	AAAGTGCTCA	TCATTGGAAA	ACGTTCTCTG	1920
GGGCGAAAAA	TCTCAAGGAT	ATCTTTTACT	TTGAGATCCA	GTTCCGATGA	ACCCACTCGT	1980
GCACCCAACT	GATCTTTCAG	ATCTTTTACT	TTTCTGGGTT	AGCAAAAAAC		2040
GGAGGCAAAA	ATGCCGCAAA	AAAGGGAATA	AGGGCGACAC	GGAAATGTGT	AATACTCATA	2100
CTCTTCCTTT	TTCAATATTA	TTGAAGCATT	TATCAGGGTT	ATTGCTCTAT	GAGCCGATAC	2160
ATATTGGAAT	GTATTTAGAA	AAATAAACAA	ATAGGGGTTT	CGCGCACATT	TGCCCGAAAA	2220
GTGCCACCTA	AATTGTAAGC	GTTAATATTT	TGTTAAATTT	CGCGTTAAAT	TTTTGTTAAA	2280
TGAGCTCATT	TTTTTAACCA	TAGGCCGAAA	TCGGCAAAAT	CCCTTATAAA	TCAAAAGAAT	2340

AGACCGAGAT	AGGGTTGAGT	GTGTTCCAG	TTTGGAACAA	GAGTCCACTA	TTAAGAAGC	2400
TGGACTCCAA	CGTCAAAGGG	CGAAAAACCG	TCTATCAGG	CGATGGCCCA	CTACGTGAAC	2460
CATCACCCTA	ATCAAGTTTT	TTGGGGTCGA	GGTGCCGTAA	AGCACTAAAT	CGGAACCCCTA	2520
AAGGGAGCCC	CCGATTTAGA	GCTTGACGGG	GAAAGCCGGC	GAACGTGGCG	AGAAAGGAAG	2580
GGAAGAAAGC	GAAAGGAGCG	GGCGCTAGGG	CGCTGGCAAG	TGTAGCGGTC	ACGCTGCGCG	2640
TAACCCACCAC	ACCCGCCGCG	CTTAATGCGC	CGCTACAGGG	CGCGTCCCAT	TCGCCATTCA	2700
GGCTGCGCAA	CTGTGTGGAA	GGGCGATCGG	TGCGGGCCTC	TTCCGTATTA	CGCCAGCTGG	2760
CGAAAGGGGG	ATGTGCTGCA	AGGCGATTAA	GTGGGTAAAC	GCCAGGGTTT	TCCCAGTCAC	2820
GACGTGTGTA	AACGACGGCC	AGTGAGCGCG	CGTAATACGA	CTCACTATAG	GGCGAATTGG	2880
AGCTCCACCG	CGGTGGCGCG	CGCTCTAGAA	CTAGTGGATC	CGTCGACTAG	AGGGCCCCGAC	2940
GTCGAACCTA	GGCACTAAGG	GATGTGAGGC	CAGCATCACC	GTTCGAGAAA	TTGACACAAG	3000
CTCACCACCA	ATTTTCCCAA	TAGAGTTTCA	TTTCTTCGTC	GTCAGCAGCT	GGCTGTGACCA	3060
TGTAGTCACA	CATGGAAGCG	CTACACCCCA	AGTTGCAATA	CTTGACGGTG	TCTGTGTTACT	3120
CTGAGTTGGA	CACAAAGGGC	AAATTGGGGA	AGCCTTTCGG	GCATTTTCGG	CTACTAGTCA	3180
GCTTACACTT	GCAGACGCTT	CGCGAAAGCT	TCTTGGCGCC	TTTGACTTTG	CAAAGGTTGT	3240
AGCACTTCTT	TCCCAAGGTA	CTCTTGACG	AACTCTTGCC	TTTACTTTGC	ACCTGTTTCA	3300
GAACCAACCC	CAGTATAAGT	AAACACACCA	TCACACCCTT	GAGGCCCTTG	CTGGTGGCCA	3360
TGGTG						3365

(2) INFORMATION FOR SEQ ID NO:3:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 5360 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

CTAAATTGTA	AGCGTTAATA	TTTTGTTAA	ATTCGCGTTA	AATTTTGT	AAATCAGCTC	60
ATTTTAAAC	CAATAGGCCG	AAATCGGCCA	AATCCCTTAT	AAATCAAAG	AATAGACCGA	120
GATAGGGTGT	AGTGTGTTC	CAGTTTGGA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGTCAAA	GGGCGAAAA	CCGTCTATCA	GGCGGATGGC	CCACTACGTT	AACCATCACC	240
CTAATCAAGT	TTTTTGGGTT	CGAGGTGCGC	TAAAGCACTA	AATCGGAAAC	CTAAGGGGAG	300
CCCGCGATT	AGCGCTTGAC	GGGGAAGGCC	GGCGAACGTT	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	GGGGGCGCTA	GGGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACCCAC	420
CACACCCGCC	CGCGCTTAATG	CGCGCTACAC	GGGCGCGTCC	CATTCCGCAT	TCAGGCTGGG	480
CAACTGTGGG	GAAAGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGGAAGG	540
GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCAGT	CACGACGTTG	600
TAAACGACG	GCGACTGAGC	CGCGGTAAAT	CGACTCACTA	TAGGGCGAAT	TGGAGCTCCA	660
CCCGGCGTGC	GGGCCGCTCA	GATTATATA	TTTATAAGCT	AAACAAACCCG	GCCCTAAAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TGCTCGCACG	780
GAATTGCATG	TTTCTTGTTG	GAAACATATT	CACGCAATCT	CCACACATAA	AGGTTTATGT	840
ATAAACCATT	ATTAGCTTCA	GTTTAAATTAC	AGTCTTATTT	GGATGCAATAT	GTATGGTTCT	900
CAATCCCATAT	AAGTTAGAGT	AAAAATAAG	TTTAAATTTT	ATCTTAATTC	ACTCCAACAT	960
ATTATGGATCT	ACAATACTCA	TGTGATCCCA	AACAACTTAC	TTATATTGAG	GTGAATTTGG	1020
TAGAAATTAA	ACTAAGCTAC	ACACTAAGCC	AATCTTACT	ATATTAAAGC	ACCAAGTTTCA	1080
ACGATGCGTC	CGGCTCAATA	TTATTAAAAA	ACTCCTACAT	TTCTTTATAA	TCACCCCGCA	1140
CTCTTTAATG	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGTGTGAA	1200
TTATCTATAA	GTGTTCTGGA	TATTGGTTGT	TGGCTCCCAT	ATTCACACAA	CAATATCAAT	1260
AGAAAAACATA	TGTTTTTATG	AAACAAAATT	TATCATATAT	CATATATATA	TATATATCAT	1320
ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATAACTA	GTGCAACTTA	ATACATGTGT	1380
GATTATAGAG	GAAATAAGAGG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGATCGGAA	1440
AGGGGTTGGA	AACGATTAAA	CGATTAAATC	TCTTCTTAGT	CAAAATTTGAA	TAGAAGGAGA	1500
TTTAATATAT	CCCAATCCCC	TTCGATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAGTGGT	1560
GAGGAACACG	AAAGAACCAT	GCATTGGCAT	GTAAGCTCC	AAGAATTTGT	TGTATCCTTA	1620

ACAACTCACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAG	AAACAATCAA	1680
ACAAATCCTC	TCTGTGTGCA	AAGAAACACG	GTGAGTCATG	CCGAGATCAT	ACTCATCTGA	1740
TATCATGTCT	TACAGCTCAC	AAGACATTAC	AAACAACATCA	TATTGCATTA	CAAGATCTGT	1800
TTCTAGTAAA	ATAAATAGG	CCGGACAGGA	CAAAAAATCCT	TGACGTGTAA	AGTAAATTTA	1860
CAACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCTG	TTTTACGTAG	ATCAACAACC	1920
TGTAGAAAGC	AACAATACTG	AGCCACGCAG	AAGTACAGAA	TGATTCCGAG	TGAACCATCG	1980
ACGTGTCTAG	TAAAGAGAGT	GACGAGTCAT	ATACATTGCG	CAAGAAACCA	TGAAGCTGCC	2040
TACAGCGCTG	TCGGTGGCAT	AAGAACAACA	GAATTTGTGT	TAATTAATCA	AAAGTATATA	2100
TAACGCTCGG	ATGCCTGTGC	ACTTCTCCAT	CACCACCATT	GGGTCTTCAG	ACCATTAGCT	2160
TTATCTACTC	CAGAGCGCAG	AAGAACCCTGA	TCGACACCAT	GGCCACCAGC	AAGGCCCTCA	2220
AGGGTGTGAT	GGTGTGTTTA	CTTATACTGG	GGTGGTTCT	CGAACAGGTG	CAAGTAGAAG	2280
CGCAAGAGTTG	CTGCAAGAGT	ACCCTGGGAA	GGAAAGTGCTA	CAACCTTTGC	AAAGTCAAGG	2340
CGCCCAAGAA	GCTTTGCGCA	GGCGTCTGCA	AGTGTAAGCT	GACTTAGTAGC	GGAAAAATGCC	2400
ATTGCAACTT	GGGGTGTAGT	GCTTCTCATGT	GTGACTACAT	TGAAACAGAG	ACCGTCAAGT	2460
ACGAAGAAAT	GAACCTCTAT	TTGGAAAAAT	GTGGTGATGC	TTGTGTCAAT	TTCTGCAACG	2520
GTGATGCTGG	CCTCACATCT	CTTAGTGCCT	AAGTGTGCAG	TCGGGCCCTC	TAGTCCAGCG	2580
ATCCCGCGCG	GTGTCCCTCC	CTGAAGAAAC	TATGTGCTGT	AGTATAGCCG	CTGCCCGCTG	2640
GCTAGCTAGC	TAGTTGAGTC	ATTTAGCGGC	GATGATTGAG	TAATAAGTGT	TCAGCATACA	2700
CCATGTCATG	GTGGCAGTGT	CAGTGTGAGC	AATGACCTGA	ATAGATTTAT	GAATATGAAA	2760
GAAAAAAGTA	TTGTTCCAAA	TTAAACGTTT	TAACCTTTTA	GTATATATTG	GGATATATGC	2820
ATATATGTTT	TCGTGATATG	GCTAATTGTA	AATGAAGGGA	AGTTAAATGA	TACAGTGATAC	2880
ATCTAAGAAC	TAAACAAAGT	CACCGACATC	CGTAGGGCTA	CCCAACAATT	TTGATCGACT	2940
GATGAGATCC	CTCGTAATAT	CGTAGGGCTA	CCCGGTACCC	AGCTTTTGTG	CCCTTTAGTG	3000
GTGTATTAC	ATTGTTTGGC	GAGTGGGGGG	GTCTATAGCT	TTTCTGTGTG	GAATTTGTTA	3060
CAACGGAAGC	GAGTCGACCT	CGTAATCATG	CGGAAAGCTG	CGCTCGGTGC	CTCGGGGTGC	3120
ACGGGTTAAT	TTTTCCTGCG	CGTATCGTGT	TTTCTGTGTG	TTTCTGTGTG	TTTCTGTGTG	3180
TTCCGTCACA	ATTCCACACA	ACATACGAGC	CGGAAGCTAA	AAGTGTAAAG	TCAGCTCGGG	3240
CTAATGAGTG	AGCTAACTCA	CATTAAATG	GTTCGCCTCA	CGCCCGCTCT	TCCAGTCCGG	3300
AAACCTGTGC	TGCCACGTGC	ATTAATGAAT	CGGCCAACGC	CGCGGGAGAG	GGCGGTTTGC	3360
TATTTGGGCG	TCTTCGCTGT	CCTCGCTCAC	AATACGGTTA	TCCACAGAA	CAGGGGATGA	3420
CGCAGCGGTA	TCAGCTCACT	CAAAAGCGGT	GCAAAGAGCC	AGGAACCGCTA	AAAAGCGGCG	3480
CGCAGGAAAG	AACATGTGAG	CAAAAGGCCA	CCCTGACGAG	CATCACAATA	ATCGACGCTC	3540
GGTCTGCGCG	TTTTTCCATA	GGCTCCGCC	ATAAAGATAC	CAGGCGTTTC	CCCTGGGAG	3600
AAGTCAGAGG	TGGCGAAACC	CGACAGGACT	GCCGCTTACC	GGATACCTGT	CCGCTTTCTG	3660
CTCCCTCGTG	CGCTTCTCTG	TTCCGACCT	CTCAGCTGT	AGGTATCTCA	GTTCGGTGTG	3720
CCCTTCGGGA	AGCGTGGCGC	TTTCTCATAG	CGAACCCCC	GTTCAGCCCG	ACCCTGCGCG	3780
GGTCTGTGCG	TCCAAGCTG	GCTGTGTGCA	CCCGTTAAGA	CACGACTATT	CGCCACTGGC	3840
CTTATCCGGT	AACTATCGTC	TTAGTCAGAG	GAGGTATGTA	GGCGGTGCTA	CAGAGTTCTT	3900
AGCAGCCACT	GGTAACAGGA	GCTACACTAG	AAGGACAGTA	TTTGTATCTG	CGCCTCTGCT	3960
GAGTGGTGG	CCTAACTACG	GCTACACTAG	TAGCTCTTGA	TCCGCGAAAC	AGGACTCCGC	4020
GAAGCCAGTT	ACCTTCGGAA	AAAGAGTTGG	GCAGATTACG	CGCAGAAAAA	AAAGACTCTA	4080
TGGTAGCGGT	GGTTTTTTTG	TTTCAAGCA	TGACGCTCAG	TGGAACGAAA	ACTACAGTTA	4140
AGAAGATCCT	TTGATCTTTT	CTACGGGCTG	GATCTTCACC	TAGATCTCTT	TAATATATAA	4200
AGGGATTTTG	GTCACTGAGT	TATCAAAAAG	TGAGTAAACT	TGGTCTGACA	GTTCACCAATG	4260
ATGAAGTTTT	AAATCAATCT	TTCTACGGAT	CTGTCTATTT	GGTCTGCTCA	TAGTTGCGTG	4320
CTTAATCAGT	GAGGACACTA	CTACGATACG	GAAGGGCTTA	CCATCTGAAA	CCAGGTGCTG	4380
ACTCCCGCTC	GTGTAGATAA	GCTCACCCTG	TCCAGATTTA	TCAGAGTGGC	ACAGCCGACG	4440
AATGATACCG	CGAGACCCAC	GTCGACCGGC	AACTTTATCT	GGCTCCATCC	AGTCTATTAA	4500
CGGAAGGGCC	GAGCGCAGAA	GTGGTCTGCG	GCCAGTTAAT	AGTTTGGCGA	ACGTTGTTGG	4560
TTGTTTCCGG	GAAGCTAGAG	TAAGTAGTTC	GTGCTTTGGT	ATGGCTTCTC	TCAGCTCCCG	4620
CATTGTCTCA	GGCATCGTGG	TTACATGATC	CCCCATGTTG	TGCAAAAAAG	CGGTTAGTCT	4680
TTCCCAACGA	TCAAGCGGAG	TCAGAGATGA	GTGGCCGCTA	GTGTTATCAC	TCATGTTTAT	4740
CTTCGGTCTC	CCGATCGTTG	TCAGAGATGA	GCCATCCGTA	AGATGCTTTT	CTGTGACTGG	4800
GGCAGCACTG	CATAATTCTC	TTACTGTGAT	GTGTATGCGG	GCACCGACTG	GTCTTTGGCC	4860
TGAGTACTCA	ACCAAGTCAT	TCTGAGAATA	TAGCAGAACT	TTAAAAAGTG	TCATCATTTG	4920
GGCGTCAATA	CGGGATAATA	CCGCGCCACA	GATCTTACCG	CTGTTGAGAT	CCAGTTCGAT	4980
AAACGCTTCT	TCGGGGCGAA	AACCTCTCAAG				5040
						5100

GTAACCCACT CGTGCACCCA ACTGATCTTC AGCATCTTTT ACTTTCACCA GCGTTTCTGG 5160
 GTGAGCAAAA ACAGGAAGGC AAAATGCCGC AAAAAGGGGA ATAAGGGCGA CACGGAAATG 5220
 TTGAATACTC ATACTCTTCC TTTTTCATA TTATTGAAGC ATTTATCAGG GTTATTGTCT 5280
 CATGAGCGGA TACATATTGT AATGTATTTA GAAAAATAAA CAAATAGGGG TTCCGCGCAC 5340
 ATTTCCCCGA AAGTGCCAC

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5511 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

TCGCGCGGTT CGGTGATGAC GGTGAAAACC TCTGACACAT GCAGTCCCCG GAGACGGTCA 60
 CAGCTTGTCT GTAAGCGGAT GCCGGGAGCA GACAAGCCCG TCAGGCGGGTG 120
 TTGCGGGGTG TCGGGGCTGG CTTAACTATG CGGCATCAGA GCAGATTGTA CTGAGATGC 180
 ACCATATCGG GTGTGAAATA CCGCACAGAT GCGTAAGGAG AAAATACCGC ATCAGGCGCC 240
 ATTGCGCATT CAGGCTGCGC AACTGTTGGG AAGGCGGATC GGTGCGGGCC TCCTCGCTAT 300
 TACGCCAGCT GGGGAAAGGG GGATGTGCTG CAAGGCGATT AAGTTGGGTA ACGCCAGGGT 360
 TTCCCGAGT ACAGCTTTGT AAAACGACGG CCAGTGAATT CTTTATGAA TAATAAAT 420
 GCATATCTGT GCATTACTAC CTGGGATACA AGGGCTTCTC CGCCATAACA AATTGAGTTG 480
 CGATGCTGAG AACGAACGGG GAAGAAGTA AGCGCGCCCC AATGAAAAGG GAACGATATG 540
 TCGGCTATAG CAGGTGAAAG TTCGTGCGCC CATACTAA TCCACTAAAG TGTCTATCT 600
 TTGGGATACT TAAATTTGGA ATATAAGATT TATATCTAAT CTGTTGGAGT TGCTTTTAGA 660
 TTTTAAACAG TCTAGGACGG ATATAAGATT TATATCTAAT CTGTTGGAGT TGCTTTTAGA 720
 GTAACCTTTT TCTCTGTTTC GTTTATAGCC GATTAGCACA AAATTAACCT AGGTGACGAG 780
 AAATAAGAA AAACGGAGCG AGTAAAAAAT ACCCAAAAAA TACTTTGGAG ATTTTGTCT 840
 CAAAATATC TTCTAATTT AAAAGCTACA TATTAATAAT ACTATATATT AAAAATACCT 900
 CGAGATCACT GCTTGGGATG GGCAGGGCCA ATAGCTAATT GCTAAGGATG GGTATATATT 960
 ATGTATCGTG TGAACATAT AGGGGCTAAT AGTTAGATGA ATACAAGAAC TCACTTTTGG 1020
 GGGGTGCTGT TTGAGCCTAG CGATGAAGGG TCATAGTTTC GATGCCAAAC TCCGCAAGGG 1080
 TTCGTCTGCT GTGCTGTTT TCAGCGTAAC GGCATCAATG TCATTATTTA TTTATGGACT 1140
 GACAAATGAA GAAGCGAAGA GATTATAGAA ACGTACATAC TACAGAGCCA CACTTATTGC 1200
 TGCTCAGTA GCTTACAGCA TCGTACCCGC ACGTACATAC TATATACATA CAGCTCCAAC 1260
 ACTGCTCCGC GCTTACGAT ATAGTTAAAC CGTCCGTCGC CGGCCACCTC CTGTTGCTT 1320
 CTTCCACTC AGGCTCATGC TACGTACGCA ACGTCAAGCT TAGGCCACTAA GATATCTGAG 1380
 GCTGCTCGTT TTGGCGAGCT AGAGGGCCCG ACGTCAAGCT TAGGCCACTAA AATAGAGTTT 1440
 GCGAGCATCA CGGTTGCAGA AATTGACACA AGCATCACCA CATGTAGTCA CACATGGAAG 1500
 CATTTCTTCG TCGTCAGCAG CTGCGTTGAC CATGTAGTCA CACATGGAAG CCCTACACCC 1560
 CAAGTTGCAA TACTTGACGG TGCTCTGGTTG TGTCTGTTG GACACAAGGG CCAATTTGGG 1620
 GAAGCCTTTC GGGCATTTTC CGCTACTAGT CAGCTTACAC TTGCGAGCGC CTGCGCAAGG 1680
 CTTCTGGGCG CTTTGAAGTT TGCAAAAGGTT GTAGCACTTC TTTCCACGGG TACTCTTGCA 1740
 GCAACTCTTG CTTTCTACTT GCACCTGTTC CATGGTGTAG TGTCGACTGT GGCAGCGATT 1800
 CATCACACCC TTGAGGCCCT TGCTGGTGGC CATGGTGTAG TGTCGACTGT GGCAGCGATT 1860
 GATGTGTTGT GGATCCTTGG GTTGCTGTGA TGCAGAACTA AAGCGGAGGT GGAAGGCTGC 1920
 TTATCACAGC CCGGGCCCTG GTACGTGGCG CGGCCGCGCG GCTACGTGGA GGAAGGCTGC 1980
 GTGGCAGCAG ACACACGGGT CGCCACGTCC AGGGTGTGGC CGCCTCTGAG CAGACTTTGT 2040
 GCTCCGCGCT GGTGCACGCC GGTGCTGAGC GCGCGGAGCG ACTCCGATCC TAAAGAGATT 2100
 AGTGGTGTGC TGTTCCGGGG ACTCCGATCC ACGGGCGAGC GACAAATTTG TGAAGAGAT 2160
 CCTACTAGGT ACGTTCATTG TATCTGGACG ACGGGCGAGC TAAATTTGAT TCAACAAATA 2220
 GGGCAGTTTT TTTTGTAGAA AACAGAGAAT TCCGTTGAGC TAATTTGAT TCAACAAATA 2280
 AGCTATTAG TGGTTTAGC TTAGATTAAG GAAGCTAACG ACTAATAGCT AATAATTAGT 2340
 TGGTCTATTA GTTGACTCAT TTTAAGGCC TGTTTCAATC TCGCGAGATA AACTTTAGCA 2400

GCTATTTTT	AGCTACTTTT	AGCCATTGTT	AATCTAAACA	GGAGAGCTAA	TGGTGGTAAT	2460
TGAAACTAAA	CTTTAGCATT	TCAATTCATA	TAGCTAAAGT	TTAGCAGGAA	GCTAAACTTT	2520
ATCCCCGTGAG	ATTGAAACGG	GGCCTAAATC	TCTCAGCTAT	TTTTGATGCA	AATTACTGTC	2580
ACTACTGGAA	TCGAGCGGTT	TGCCGAGTGT	CAAAGCCTGA	AAAACACTCC	GTAAGAGACT	2640
TGCTAGTGT	GACACTCGAC	AAAGAGATCT	CGACGAACAG	TACATCGACA	ACGGCTTCTT	2700
TGTCGAGTAC	TTTTTATCGG	ACACTTGACA	AAGCTTTTGT	CGAGTGAAC	ACATTGAAAC	2760
TCTATGATT	TATGTGTAGG	TCACTTAGGT	TTCTACACAT	AGTACGTCAC	AACTTTACCG	2820
AAACATTATC	AAATTTTTAT	CACAACCTCT	ATATATGATA	TCATGACATG	TGGACAAGTT	2880
TCATTAATTT	TCGACTTTAT	TTGTGTTTTA	TACAATTTTT	AAACAACATG	ATAACAAGTT	2940
CACGGTCATG	TTTAGTGAGC	ATGGTGGCTG	AAGATTTCTG	TCTGCTTCTG	AAATCGGTCG	3000
TAACCTTGTC	TAGATAACAT	GCATATCATT	TATTTTGCAT	GCACGGTTTT	CCATGTTTCG	3060
AGTGACTTGC	AGTTTAAATG	TGAATTTTCC	GAAGAAATTC	AAATAAACGA	ACTAAATCTA	3120
ATATTTATAG	AAAAATTGTT	TGTAATATG	TAATTTTGCC	AAAATGGTAT	ATGTAGATCT	3180
ACATAGTGTA	GGAACATACC	ACAAAAAGTT	TGGTTGGCAA	AATAAAAAAA	ATAAAAATATA	3240
CTTTATCGAG	TGTCACAAGGA	TGGCACTCGG	CAAGCTTGCC	GTAATCATGG	TCATAGCTGT	3300
TTCTGTGTG	AAATTTGTTT	CCGCTCACAA	TTCCACACAA	CATACGACGG	GGAAAGCATAA	3360
AGTGTAAAGC	CTGGGGTGCC	TAATGAGTGA	GCTAACTCAC	TTAATGATCT	GGCCACCGGG	3420
TGCCCCGCTT	CCAGTCGGGA	AACCTGTCGT	GCCAGCTGCA	TTAATGATCT	GACTCGCTGC	3480
CGGGGAGAGG	CGGTTTGGCT	ATTGGGCGCT	CTTCCGCTTC	CTCCGCTCACT	ATACGGTTAT	3540
GCTCGGTGCT	TGCGCTGCGG	CGAGCGGTAT	CACTCTCACT	AAAGGCGGTA	ATACGGTTAT	3600
CCACAGAATC	AGGGGATTAAC	CGAGGAAAGA	ACATGTGAGC	AAAGAGCCCA	CAAAAGGCTCA	3660
GGAAACGATA	AAAGGCGCGG	TTGCTGGCGT	TTTTTCCATG	GCTCCGCCCC	CCTGACGAGC	3720
ATCACAAAAA	TGACAGCTCA	AGTCAGAGGT	GGCGAAACCC	GACAGGACTA	TAAGATGATC	3780
AGGCGTTTCC	CCCTGGAAGC	TCCTCTGTCG	GCTCTCCTGT	TCGAGCCCTG	CCGCTTACCG	3840
GATACCTGTC	CGCTTTTCTC	CCTTCGGGAA	CTTCTCAATG	TCACGCTGTA	TAAGATGATC	3900
GGTATCTCAG	TTGCGTGTAG	GTCGTTGCTC	CCAAGCTGGG	CTGTGTGAC	GAACCCCCCG	3960
TTACGCCCGA	CCGCTGACCC	TTATCCGGTA	TGAGTCCAAC	CCGCTGAAGC	AGGTATGTAG	4020
ACGACTTTATC	GCCACTGGCA	CGAGCCACTG	GTAACAGGAT	TAGCAGAGCG	AGGTATGTAG	4080
CGCGTGCTAC	AGAGTCTCTG	AAGTGGTGGC	CTAACACCGG	CTACACTAGA	AGGACAGTAT	4140
TTGGTATCTG	CGCTCTGCTG	AAGCCAGTGA	CCTTCGGAAA	AAGAGTTGGT	AGCTCTTGAT	4200
CCGGCAAAAC	AACCCAGCGT	GGTAGCGGTG	GTTTTTTTTT	TTGCAAGCAG	CAGATTACGC	4260
GCAGAAAAAA	AGGATCTCAA	GAAGATCCTT	TGATCTTTTC	TACGGGGTCT	GAGCTCAAGT	4320
GGAAACGAAA	CTCACGTTAA	GGGATTTTGG	TCATGAGATT	ATCAAAAAGG	ATCTTCACCT	4380
AGATCCTTTT	AAATTAATAA	TGAAGTTTTA	AATCAATCTA	AGGATATATAT	GAGTAAACCT	4440
GGCTGACAGC	TTACCAATGC	TTAATCAGTG	AGGCACCTAT	CTCAGCGCAT	TGCTATTTTC	4500
GTTTCATCAT	AGTTGCCTGA	ATGATACCGC	GAGACCCACG	TACGATACCG	GAGGCGTTAC	4560
CATCTGGCCC	CAGTGCTGCA	ATGATACCGC	GAGACCCACG	CTCAGCGCAT	CCAGATTTAT	4620
CAGCAATAAA	CCAGCCAGCG	GGAAAGGGCG	AGCGCAGAGG	TGGTCTTGCA	ACTTTATCCG	4680
CTCCATCCCA	GTCTATTAAAT	TGTTGCCGCG	AAGCTAGAGT	AGTATAGTTC	CCAGTTAAATA	4740
GTTTGCAGCA	CGTTGTTGCC	ATTGCTACAG	GCATCGTGGT	TACATGATCC	CCCATGTTGT	4800
TGCCATTCATT	AGCTTCCGGT	TCCCAACGAT	CAAGGCGAGT	CAGAAGTAAG	TGTGGCCGAG	4860
CGAAAAAAGC	GGTTAGCTCC	TTCCGTCCTC	CGATCGTTGT	TACTGTCTATG	CCATCCGTA	4920
TGTTATCACT	CATGTTTATG	CGACGACTGC	ATAATTCTCT	CTGAGAATAG	CCATCCGTA	4980
GAGCTTTTTC	TGTGACTGGT	GAGTACTCAA	CGGATAATAC	GGGATAATAC	TGATGTGGCG	5040
GACCGAGTTG	CTCTTGCCCG	CGCTCAATAC	CGGGGCGAAA	ACTCTCAAGG	ATCTTACCGC	5100
TAAAGATGCT	CATCATTTGA	AAACGTTCTT	CGGGGCGAAA	ACTCTCAAGG	ATCTTACCGC	5160
TGTTGAGATC	CAGTTCGATG	TAACCCACTC	GTGCACCCAA	GCATCTTTCA	GCATCTTTTA	5220
CTTTCACCGC	CGTTTCTGGG	TGAGCAAAAA	CAGGAAGGCA	AAATGCGCGA	AAAAAGGGAA	5280
TAAGGGCGAC	ACGGAATGCT	TGAATACTCA	TACTCTTCTT	TTTTCAATAT	TATTGAAGCA	5340
TTTATCAGGG	TTATTGTCTC	ATGAGCGGAT	ACATATTGTA	ATGTATTTAG	AAAAATAAAC	5400
AAATAGGGGT	TCCGCGCACA	TTTCCCGGAA	AAGTGCACCC	TGACGCTCAA	GAAACCATTA	5460
TTATCATGAC	ATTAACTCAT	AAAAATAGGC	GTATCACGAG	GCCCTTTCTG	G	5511

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 5115 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

GTGGGAGCT	CTCCCCATATG	GTGACCTGTC	AGGCGGCCGC	TCTAGAACTA	GTGGATCCCC	60
CCCTCGAGGT	CGACGGTATC	GATAAGCTTG	ATATCTTACA	AGGCCACGCC	CAGCGACCTA	120
TTACACAGCG	CGCTCGGGCG	CGCGACGTCG	GGACACATCT	TCTTCCCCCT	TTTGGTGAAG	180
CTCTGCTCGC	AGCTGTGGCC	CTCCTTGGAC	GTTCGTGTGG	CAGATTCTAT	TGTTGTCTCG	240
TCCTCTGTGC	TTCTTGGGTA	GCTTGTGTAG	TGGAGCTGAC	ATGGTCTCAT	CAGCTTTAAA	300
ATTTGCTCT	AGACGAGGAG	TACCAGCACA	GCACGTTGCG	GATTTCTCTG	CCTGTGAAGT	360
GCACGCTCTA	GGATTGTTCAC	ACGCCTTGGT	TGGGCGGCC	TCCGTGTCTT	CGTGTACAGT	420
GTGAGCAGAG	CAGCAACAGC	CAGCAGAGAG	CGGAGAGCGA	GCCGTGACAG	GGGAGGTGGT	480
ACGCGCGCGC	CGGCGACACG	CAGCGACAGG	CGCCTCGCGT	CGCGTCCGGT	CGATGCGGGT	540
GTGGAAGTGG	AGCCGCGCGC	CCGCGCCGCC	TCCGGAACGC	AACGTTGGCT	CGTGTACAGT	600
GACAAGCGAA	GGCGCCAAAG	CGATCCAAAG	CCGCTGGGCG	ACGCGTCTAT	GGACGGACCG	660
AGAGCCAGCG	ACAAGCAGCC	GAGAAGCGAA	CGGCGTGTGC	GTGCGTGCAG	ACGACAAGCC	720
GGGCGACGCT	TCCAACGGGG	CCACGTACGC	TTTTGGGCGC	GAGCGCTGGC	GTGCGGGTCA	780
AAGCGGAGGC	AGCCCCCGAT	CGGGAAGCG	GTATCGTGGG	GGGCGCGGGG	GGAGGAGAGC	840
GTGCGTGGTG	CGCAGTGGCG	GGGGAACGG	GGTCAACGAA	CGCGCCCGCC	GACTGCGCCT	900
GTGCGGAGG	CGGAGAGCAG	CGCGCGCGCG	CTCGGACCGG	GGGGGGGCCC	CGTCACATCC	960
CCCCCTCCGC	CGCGCGTAGA	AATACCGAGG	CACCACCCCG	CGAGCGCAGC	CGACAGCCGC	1020
ATCCATCGAC	CGATCGATCG	CCACAGCCAA	AGTGAAGGGG	GAGAAGTGTA	CTGCTCCGTC	1080
CAGGAGGAGC	GAATAAACCTC	ACTGCCAGCC	CTCATCTCGT	CGACGACCAG	GTCTGTGTTT	1140
GACCACTGCG	CGCACCGGCC	GGCAGGGCTG	TCCAGATCCT	GGGCGGTATC	TGCGTGTGTC	1200
GATCCGATCC	GATCCTGTCT	TTGAGTTTCG	CCGTGACAC	GTCTTTTCTC	TCTCTCCTAC	1260
ATGATCCAGC	TTCTTCGAAC	CTAAATCTGT	AGGGCCTCAA	GGGTGTGATG	GTGTGTTTAC	1320
CGAGTGGATT	AGTCGCCATG	GCCACCAGCA	AAGTAGAAGG	CAAGAGTTGC	TCGAAGAGTA	1380
TTTACTTGGG	GTGGTTCTCT	GAACAGGTGC	AACTCAAAGG	CGCCAAAGAG	CTTTGCGCAG	1440
CCCTGGGAAG	GAAGTGCTAC	AACTTTTGCA	GAATGCTGCC	GAAGGGCTTC	CCCAAATTTG	1500
CGCTCTGCAA	GTGTAAGCTG	ACTAGTAGCG	CGTCAAGTA	TTGCAACTTG	GGGTGTAGGG	1560
CCCTTGTGTC	CAACTCAGAT	GAACCAGACA	CTGTCAGACA	CGAAGAAATG	AAACTCTATT	1620
CTTCCATGTG	TGACTACATG	GTCAACGCAG	TCTGCAACGG	TGATGCTGGC	CTCACATCCC	1680
TGGAAAAATT	TGGTGATGCT	TGTGTCAATT	AGATGCGGCC	CGGGTGAAGA	GTTCGCCCTG	1740
TTAGTGCCCTA	AGTTCGACGT	CGGGCCCTCT	GTGATGCAAA	GATGTTGGGA	CATCTTCTTA	1800
CAGGGCCCCCT	GATCTCGCGC	AAGTATGTGT	AGATGCTTGC	TTGTGCTAGT	GTAATGTAGT	1860
TTGCTTTATG	TGATATGGAC	AACCTAATAA	CGCATGAAC	TAATTGCTTG	CGTGTGTAGT	1920
GTAGTGGTGG	CCAGTGCAC	TATATTGCGA	GTCTCGTAT	AGATATTATA	TAGGTGAGTA	1980
TAAGTACCGA	TCGGTAATTT	TATATTGCGA	CTCTCGTAT	CATTACTGTA	AAATTTCTGC	2040
AATTAATCCT	CGTGTTCGGT	GTCTTATATG	ATTCTATAT	TTCCGGCTCT	GGCAGCTTCT	2100
TTTTCTCTCT	CTGAATCCTA	CGTGTGTGAA	CCCTGTTTGT	TTCACTTATG	TCTATATAAA	2160
GTTCAAAAAG	AGACATAGC	CTATCTTTGG	CGCTCAGATT	TAATCGGTTG	AGTGTGTTGA	2220
GGCCACCAAA	AGCTGCTGCG	GACTGCCAAA	TATAAACACA	TAATCGGTTG	AGTGTGTTGA	2280
TTAGTTGGGG	CAAAAACCAT	CCAAAATCAA	GAGCCCTATG	AACAACCTTA	TCTTTCTCCA	2340
ATATTAGGAA	TCTGTCACTT	TCTAGATCCT	CCACAGCCAG	ATTCTCCTCA	CGCCAGGATT	2400
TACGTAATCG	TAATGATACT	CAGATTCTCT	CCACAGCCAG	CCTTGTGTTG	TAGAGCTTTA	2460
TTACAGAAAG	CTGGTCAGAA	AAAAGTTAAA	CCAAACAGAG	GGTTCGGATT	TAGAGCTTTA	2520
GGCTTTCCCC	CTCAAGCTCT	AAATCGGGGG	CTCCCTTAG	CACGTAGTGG	GGCATTGCCC	2580
CGCCACCTCG	ACCGCAAAAA	ACTTGATTGG	GGTGATGGTT	TCTTTAATAG	TGAGCTCTTG	2640
TGATAGACGG	TTTTTCGCC	CAACCTATC	TGCGTCTATT	CTTTTGAATT	ATAAGGGATT	2700
TTCCAAACTG	GAGCAACACT	GTTAAAAAAT	GAGCTGATT	TAAACAAAT	TAACCGGAAT	2760
TTGCCGATTT	CGGCCTATTG	TACAATTTTC	CCTGATCGCG	TATTTCTCGC	TTACGACATCT	2820
TTTAAACAAA	TATTAACGTT	TACAGTGGC	ACTTTTCGGG	GAAATGTGGG	CGGAACCCCT	2880
GTGGCGTTAT	TCACACCGCA	ATGATATCGC	TCATGAGACA	ATAACCTGGA	CCGTGTCGCC	2940
ATTTGTTTAT	TTTTCTAAAT	AAAAAGGAAG	AGTATGAGTA	TTCAACATTT	CCGTGTCGCC	3000
TAAATGCTTC	AATAATATTG					3060

CTTATTCCT	TTTTGCGGC	ATTTGCGCTT	CCTGTTTTG	CTCACCCAGA	AACGCTGGTG	3120
AAAGTAAAG	ATGCTGAAGA	TCAGTTGGGT	GCACGAGTGG	GTTACATCGA	ACTGGATCTC	3180
AACAGCGGTA	AGATCCTTGA	GAGTTTTCGC	CCCCAAGAAC	GTITTCGAAT	GATGAGCACT	3240
TTTAAAGTTC	TGCTATGTCA	TACACTATTA	TCCCGTATTG	ACGCGGGGCA	AGAGCAACTC	3300
GGTCGCCGGG	CGCGGTATTC	TCAGAAATGAC	TTGGTTGAGT	ACTCACCAAT	CACAGAAAAG	3360
CATCTTACGG	ATGBCATGAC	AGTAAGAGAA	TTATGCAATG	CTGCCATAAC	CATGAGTGAT	3420
AACACTCGCG	CCAACTTACT	TCTGACAACG	ATCGGAGGAC	CGAAGGAGCT	AACCGCTTTT	3480
TTGCACAACA	TGGGGGATCA	TGTAACCTCG	CTTGATCGTT	GGGAACCCGA	GCTGAATGAA	3540
GCCATACCAA	ACGACGAGCG	TGACACCACG	ATGCCCTGAG	CAATGCCAAC	AACGTTGCGC	3600
AAACTATTAA	CTGGCGCACT	ACTTACTCTA	GCTTCCCGTG	AACATTAAT	AGACTGGATG	3660
GAGGCGGATA	AAGTTGCAGG	ACCACTTCTG	CGCTCGGCC	TTCCGGCTGG	CTGCTTTAT	3720
GCTGATAAAT	CTGGAGCCGG	TGAGCGTGGG	TCTCGCGGTA	TCATTGCGAG	ACTGGGGCCA	3780
GATGTAAGC	CTCTCCGTAT	CGTAGTTATC	TACACGACGG	GGAGTACAGC	AACATGTGAT	3840
GACCAAGTTT	ACTCATATAT	ACTTTAGATT	GATTTAAAC	TTACCTTAA	ATTTAAAGG	3900
ATCTAGGTGA	AGATCTCTTT	TGATAATCTC	ATGACCAAAA	TCCCTTAAAG	TGAGTTTTCG	4020
TTCCACTGAG	CTCGAGACCC	CGTAGAAAAA	ATCAAGGAT	CTTCTTGAGA	TCCTTTTTTT	4080
CTGCGCGTAA	TCTGCTGCTT	GCAAAACAAA	AAACCCACCG	TACCAAGCGT	GGTTTGTGTT	4140
CCGGATCAAG	AGCTACCAAC	TCTTTTTCCG	AAGSTAACCT	GCTTCAGCAG	AGCGCAGACT	4200
CCAAATACTG	TCCCTCTAGT	GTAGCCGTAG	TTAGGCCACC	ACTTCAAGAA	CTCTGTAGCA	4260
CCGCTACAT	ACCTCGCTCT	GCTAATCCTG	TTACAGTGG	CTGCTGCCAG	TGGCGATAAG	4320
TCGTGTCTTA	CGGGTTTGGG	CTCAAGACGA	TAGTTACCGG	ATAAGCGCGA	CGCGTCGGGC	4380
TGAACGGGGG	GTTCTGTGAC	ACAGCCACGC	TTGGAGCGAA	CGACCTACAC	CGAACTGAGA	4440
TACCTACAGC	GTGAGCTATG	AGAAAGCGCC	ACGCTTCCCG	AGGGAGAGAA	GGCGGACAGG	4500
TATCCGGTAA	GCGGCAGGGT	CGGAACAGGA	GAGCGCACGA	GGGAGCTTCC	AGGGGGAAC	4560
GCCTGGTATC	TTTATAGTCC	TGTCGGGTTT	CGCCACCTCT	GACTTGAGCG	CTGATTTTTG	4620
TGATGCTCTG	CATGGGGGCG	GAGCCTATCG	AAAAACGCCA	GCAACGCGGC	CTTTTACGCG	4680
TTGTGCGCCT	TTTGCTGGCC	TTTTGCTCAC	ATGTTCTTTC	CTCGGTTATC	CCCTGATTC	4740
GTGGATAACC	GTATTACCCG	CTTTGAGTGA	GCTGATACCG	CTCGCCGACG	CCGAACGACC	4800
GAGCGCAGCG	AGTCACTGAG	CGAGGAAGCG	GAAGAGCGCC	CAATACGCAA	ACCGCTCTCG	4860
CCCGCGCGTT	GGCCGATTCA	TTAATGCAAG	GGTTTCCCGA	GSTTTCCCGA	CTGGAAAGCG	4920
GGCAGTGAGC	GCAACGCAAT	TAATGTGAGT	TAGCTCACTC	ATTAGGCACC	CCAGGCTTTA	4980
CACTTTATGC	TTCCGGCTCG	TATGTTGTGT	GGAAATTGTA	CGGGATAACA	ATTTCACACA	5040
GGAAACAGCT	ATGACCATGA	TTACGCCAAG	CTATTAGGT	GACACTATAG	AATACTAAG	5100
CTATGCATCC	AACGC					5115

(2) INFORMATION FOR SEQ ID NO:6:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 5392 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

CTAAATGTGA	AGCGTTAATA	TTTTGTTAAA	ATTCGCGTTA	AAATTTTGT	AAATCAGCTC	60
ATTTTTTAAC	CANTAGGCCG	AAATCGGCAA	AATCCCTTAT	AAATCAAAG	AATAGACCGA	120
GATAGGGTGT	AGTGTGTTTC	CAGTTTGAA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGCTCAA	GGGCGAAAAA	CCGCTCTATCA	GGCGATGGC	CCACTACGTG	AACCATCACC	240
CTAATCAAGT	TTTTGGGGT	CGAGGTGCCG	TAAAGCACTA	AATCGGAACC	CTAAGGGGAG	300
CCCCCGATTT	AGAGCTTGAC	GGGGRAAGCC	GCGAAGCTG	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	GCGGGCGCTA	GGCGCTGGC	AAGTGTAGCG	GTCACGCTGC	GCGTAACACC	420
CACACGCCGC	CGCTTAAATG	CGCCGCTACA	GGGCGCGTCA	CATTCCGCAT	TCAGGCTGCG	480
CAACTGTGTG	GAAGGGCGAT	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGAAGGG	540
GGGATGTGCT	GCAAGGCGAT	TAAGTTGGGT	AAGCGCAGGG	TTTTCCAGT	CACGACGTTG	600

TAAACGACG	GCCAGTGAGC	GCAGCGTAATA	CGACTCACTA	TAGGCGCAAT	TGGAGCTCCA	660
CCGCGGTGCG	GGCCGCTCTA	GATTATATAA	TTTATAGCT	AAACAACCCG	GCCTTAAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	GGGAGTTTCG	AACGTCCACT	TGTCGCGACG	780
GAATTGCGATG	TTTCTGTGTT	GAAGCATATT	CACGCAATCT	CCACACATAT	AGGTTTATGT	840
ATAAATCTAC	ATTAGCTCA	GTTTAATTAC	AGTCTTATTT	GGATGCATAT	GTATGGTTCT	900
CAATCCATAT	AAGTTAGATG	AAAAAATAAG	TTTAAATTTT	ATCTTAATTTC	ACTCCAACAT	960
ATATGGATCT	ACAACTACTA	TGTGCATCCA	AACAACTAC	TTATATTGAG	GTGAATTTGG	1020
TAGAAATTA	ACTAAGTTAC	ACACTAAGCC	AATCTTTACT	ATATTAAGCG	ACAGTTTCCA	1080
ACGATCGTCC	CGCGTCAATA	TTATTAATAA	ACTCTACAT	TTCTTTATAA	TCAACCCGCA	1140
CTCTTATAAT	CTCTCTCTGA	CTACTATAAT	AAGAGAGTTT	ATGTACAAAA	TAAGGTGAAA	1200
TTATCTATAA	GTGTTCTGTA	TATTTGGTTG	TGGCTCCCAT	ATTCACACATA	CCTAATCAAT	1260
AGAAAACATA	TGTTTTATTA	AAACAAATTT	TATCATATAT	CATATATATA	TATATATCAT	1320
ATATATATAT	AAACCGTAGC	AATGCACGGG	CATATAACTA	GTGCAACTTA	ATACATGTGT	1380
GTATTAAGAT	GAATTAAGAG	GTATCCAAAT	AAAAAACTTG	TTGCTTACGT	ATGGAGTCGA	1440
AGGGGTTTGA	AACGATTAAA	CGATTAAATC	TCTTCCTAGT	CAAAATTGAA	TAGAAGGAGA	1500
TTTAAATATAT	CCCAATCCCC	TTCGATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTGGT	1560
GAGGACACAG	AAAGAACCAT	GCATTGGCAT	GTAAAGCTCC	AAGAATTTGT	TGTAATCTTA	1620
ACAACTCAC	GAACATACAC	CAAAATTGCA	CGTCAAGGCT	ATTGGGTAA	AAACAATCAA	1680
ACAAATCTCT	TCTGTGTGCA	AAGAAACACG	TGTAGTCTAG	CGAGATCAT	ACTCATCTGA	1740
TATACATGCT	TACAGCTCAC	AAGACATTAC	CGTCAAGGCT	TATTCATCTA	CAAGATCTGT	1800
TTTATGAAAA	ATAAATATAG	CCGACAGGGA	ATCTAATCTG	TGACGTGTAA	AGTAAATTTA	1860
CACAAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATCTG	TTTACGTGT	ATCAACAACC	1920
TGTAGAAGGC	ACAAAACTG	AGCCACGCGAT	AAGTACAGAA	TGATCCAGAA	TGAACCATCTG	1980
ACGTGCTACG	TAAAGAGAGT	AAGAACACAA	GAATTTGTGT	TAATTAATCA	AAGCTATATA	2040
TACAGCCGTA	TGCGTGGTAC	AAGAACACAA	CACCACTCAT	GGGCTCTTCG	ACCATTTGAA	2100
TACGCTCGCT	ATGCTGTGTC	ACTCTTCCAT	CGACACCAT	GACCAAGTTC	ACAATCTCTC	2160
TTATCTACTC	CAGAGCGGAG	AAGAACCCGA	CTTGACGCGC	CTCCAAATGG	CAGCACCAGC	2220
TCATCTCTCT	TCTCTCTGTC	ATCGCCACCA	TGAACCTCAC	GCCTTCGCG	AAGCACATCA	2280
AAGATAGTCT	CCGCAAGCAG	CTTAAGGGGG	ATGATGATGA	TGACGACGAC	AATCACTATT	2340
TGGAGAGAT	CCAAGCCGCG	GGCGATGACG	TACGGAAGAA	GGAAGGAAAA	GAGCAAGACG	2400
TCAGGACCAT	GGGGGGGAA	AATCACTACA	CGCCTTTGCA	CTGGCAATTTG	GGGCTCTTAA	2460
AAGAAGAA	AGGACACATG	CAGAAGTGCT	TGGAAACCA	GAGCGAGGAA	CTGGGAGGAA	2520
GCTCGCTCAT	TTCTGTGCTG	CAGAAGATAA	CGCACTTGGC	TACTATGTGC	AGGTTTGGGG	2580
AGGAGAAGAA	AAAAATGGAG	AAGGAGCTTA	ACTAAGTTGA	TCCCCGCGCG	TGCTCCCCAC	2640
CCATGATCGG	GTGCGACTTG	TCTTCCGATG	TGGCTAGCTA	GGGTGGGAGT	CTCAGTTGGA	2700
TGAAGAAACT	ATGTGCTGTA	GTATAGCCGC	CACCATGCTAT	GGGTGGGAGT	CTCAGTTGGA	2760
GCGATGATTG	AGTAAATATG	TGTCACGCTA	AAGAAAAAAG	TATTTTCCCA	ATTTAAATCT	2820
CGAATGACCT	GAATGAACAA	ATACAATAAT	TGATATATGT	TTTCTGTATA	TGCTTAATTT	2880
TTTAACTTTT	TAATAGGTTT	ATACAATAAT	AAATCTAAGA	ACTAAAAACAA	ATGCTAATTT	2940
GTTATCATCC	ATTTAGATAT	AGACGAAAAA	TCGATGAGAT	CCCTCGTAAT	ATACCCGACA	3000
GAAATGAAGG	GAGTATATAT	TGGGATAATG	ACGTGTATTC	ACATTTGTGTG	CGCGTAGGGG	3060
TCACACGTGT	CCAGTTAATG	TATCAGTGAT	GTCAACGGAA	CGGAGTCGAC	CTCCGAGGGG	3120
TACCCACATA	TTTTGATCGA	CTATCAGAAA	TGAGGGTTAA	TGCGCGCTTT	GGCGTAATCA	3180
GGCCCGGTAC	CCAGCTTTTG	TTCCCTTTAG	TATCCGCTCA	CAATTTCCCA	CACATACGGA	3240
TGGTCATAGC	TGTTTCTGTT	GTGAAATTTG	AGCTTGGGGT	GCCTAATGAG	CACATTAATT	3300
GCCGGGAAGCA	TAAAGTGTA	AGCTTGGGGT	TTTCCAGTCT	GGAAACCTGT	CGCTAATGTA	3360
CGGTTGGGCT	CAGTCCCGCG	AGGCGGTTTG	CGTATTTGGG	CGGCGAGCGG	GCTCTTCGCG	3420
ATCGGCCAAC	CGCGGGGGAG	CGTTCCGGCTG	CGGCGAGCGG	AACGACGGAA	AGAACCTGTG	3480
ACTGACTCGC	TGCGCTCGGT	ATCAGGGGAT	GCCTTGTCTG	CGTTTTTCCA	TAGGCTCCGC	3540
GTAATACGGT	TATCCACAGA	TAATAAGGCC	TCAAGTCAGA	AGCTCCCTCG	TGCGCTCTCC	3600
CAGCAAAAGG	CCAGGAACCG	AAATCGACG	AGCTCCCTCG	GGTGGCGAAA	CGGACAGGGA	3660
CCCCCTGAGC	AGCATCACCA	TCCCCCTTTT	AGCTCCCTCG	GCTTCTTCG	TGTTCCGACC	3720
CATATAAGAT	ACAGGCGGTT	GTCGCCCTTA	CTCCCTTCG	GAAGCGTGGC	GCTTCTTCAT	3780
CTCGCGCTTA	CCGATATACCT	CAGTTCCGGT	TAGTCTGTTT	GCTCCAAAGT	GGGCTGTGTT	3840
AGCTCACGCT	GTAGGTATCT	CGACCGCTGC	GCCTTATCCG	GTAACATCTG	CTTTGAGTCC	3900
CACCAACCCC	CCGTTCAGCC	GCAGCGCTGC	GCAGCAGCCA	CTGGTAACAG	GATTAGCAGA	3960
AAACCGGTAA	GACACGACTT	ATCGCCACTG	TTGAAGTGGT	GGCTTAACCTA	CGGCTACACT	4020
GCGAGGTATG	TAGCGGCTGC	TACAGAGTTC				4080

AGAAGGACAG	TATTTGGTAT	CTGCGCTCTG	CTGAAGCCAG	TTACCTTCGG	AAAAAGAGTT	4140
GGTAGCTCTT	GATCCGGCAA	ACAAACCACC	GCTGGTAGCG	GTGGTTTTTT	TGTTTGAAG	4200
CAGCAGATTA	CGCGCAGAAA	AAAAGGATCT	CAAGAAGATC	CTTTGATCTT	TCTTACGGGG	4260
TCTGACGCTC	AGTGGAAACGA	AAACTCACGT	TAAGGGATTG	TGGTTCATGAG	ATTATCAAAA	4320
AGGATCTTCA	CCTAGATCTG	TTTAAATTA	AAATGAAGTT	TTAAATCAAT	CTAAAGTATA	4380
TATGAGTAAA	TGTGCTCTCA	CAGTTACCAA	TGCTTAATCA	GTGAGGCACC	TATCTCAGCG	4440
ATCTGTCTAT	TTCTGTTTATC	CATAGTTGCC	TGACTCCCCG	TCGTGTAGAT	AACCTACGATA	4500
CGGGAGGGCT	TACCATCTGG	CCCCAGTGCT	GCAATGATAC	CGCGAGACCC	AGGCTCACC	4560
GCTCTCAGAT	TATCAGCAAT	AAACCAGCCA	GCCGGAAGGG	CCGAGCGCAG	AAGTGGTCT	4620
GCAACTTTAT	CGCGCTCCAT	CCAGTCTATT	AATTTTGGCC	GCGAAGCTAG	AGTAAGTAGT	4680
TCGCCAGTTA	ATAGTTTGGC	CAACGTTGTT	GCCATTGCTA	GAGCATCTGT	GGTGTACGCG	4740
TCGTGCTTTG	GTATGGCTTC	ATTCAGCTCC	GGTCCCCAAC	GATCCAAGCG	AGTTACATGA	4800
TCCCCCATGT	TGTCAAAAA	AGCGGTTAGC	TCCTTCGCTC	CTCCGATCTG	TGTCAGAAAT	4860
AAGTTGGCCG	CAGTGTATAT	ACTCATGGTT	ATGGCAGCAC	TGCATAATTC	TCTTACTGTG	4920
ATGCCATCCG	TAAAGATGCT	TTCTGTGACT	GCTGAGTACT	CAACCAAGTC	ATTCTGAGAA	4980
TAGTGTATCG	GGCGACCGAG	TGCTCTTGTC	CCGCGCTCAA	TACGGGATAA	TACCGCGCCA	5040
CATAGCAGAA	CTTTAAAGAT	GCTCATCATT	GGAATAACGTT	TTCTGGGGCG	AAAACTCTCA	5100
AGGATCTTAC	CGCTGTTGAG	ATCCAGTTTCG	ATGTAAACCA	CTCGTGACCC	CAACTGATCT	5160
TCAGCATCTT	TTACTTTTAC	CAGCGTTTCT	GGGTGAGCAA	AAACAGGAAG	GCAAAATGCC	5220
GCAAAAAAGG	GAATAAGGGC	GACACGGAAA	TGTTGAATAC	TCTACTCTTT	CCTTTTTCAT	5280
TATTATTGAA	GCATTTATCA	GGGTATTGTT	CTCATGAGCG	GATACATATT	TGAATGTATT	5340
TAGAAAAATA	AACAAATAGG	GTTTCCGCGC	ACATTTCCCC	GAAAAGTACC		5392

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 5173 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CTAAATTTGTA	AGCGTTAATA	TTTGTGTA	ATTCGCGTTA	AATTTTGT	AAATCAGCTC	60
ATTTTAAAC	CAATAGGCCG	AAATCCGGCA	AAATCCCTTAT	AAATCAAAAG	AATGACCCGA	120
GATAGGGTTG	AGTGTGTTCT	CAGTTTGGAA	CAAGAGTCCA	CTATTAAAGA	ACGTGGACTC	180
CAACGTCAAA	GGCGGAAAAA	CCGTCTATCA	GGCGGATGCG	CCACTACGCT	AACCATCACC	240
CTAATCAAGT	TTTTTGGGGT	CGAGGTGCGC	TAAAGGACCT	AATCGGAACC	CTAAAGGGAG	300
CCCCCGATT	AGAGCTTGAC	GGGGAAAGCC	AAAGTGTAGCG	GCGAGAAAGG	AAGGGAAGAA	360
AGCGAAAGGA	CGGGCGCTA	GGGCGCTGCG	GCGCGCGTCC	GTACGCTGCG	CGGTAACACC	420
CACACCCCGC	CGCGTTAATG	CGGTGCGGGC	CTCTTCGCTA	TTACGCCAGC	TGGCGGTGCG	480
CAACTCTTGG	GAAAGCGCAT	TAAGTTGGGT	AACGCCAGGG	TTTTCCCGAT	TGGCGAAAGG	540
GGGATGTGCT	GCAAGCGCAT	GCGCGTAATA	CGACTCACTA	TAGGGCGAAT	TGGAGCTTCA	600
TAAACGACGC	GCCAGTGAGC	GCGCGTATAT	TTTATAAGCT	AAACAACTAC	GCCCTAAAGC	660
CGCGGGTGGC	GGCCGCTCTA	GATTATATAT	GGGAGTTTCG	AACGCTCACT	TGCTCGCAGC	720
ACTATCGTAT	CACCTATCTA	AATAAGTCAC	CAACGCAATC	CCACGATATA	AGGTTTATGT	780
GAAATGCAAT	TTTCTTGTTG	GAAAGCATAT	CAACGCAATC	GGATGCATAT	GTATGGTTCT	840
ATAAACTTAC	ATTTAGCTCA	GTTTAAATTA	AGTCTTATTT	ATCTTAATTC	ATCTCAACAT	900
CAATCCATAT	AAGTTAGAGT	AAAAAATAAG	TTTAAATTTT	TTATATTGAG	GTGAATTTGG	960
ATATGGATCT	ACATACTACT	TGTGCATCCA	AACAAACTAC	ATATTAAGCG	ACCAGTTTCA	1020
TAGAAATTA	ACTAACTTAC	ACACTAAGCC	AATCTTTACT	TTCTTTATAT	TCAACCCGCA	1080
ACGATCGTCC	CGCGTCAATA	TTATTAATAA	ACTCTACATC	ATGTACACAA	TAAAGTTGAA	1140
CTCTTATAAT	CTCTTCTCTA	CTACTATAAT	AAGAGAGTTT	ATGCACACAA	CCTAATCAAT	1200
TTATCTATAA	GTGTTCTGGA	TATTGGTTGT	TGGCTCCCAT	CATATATATA	TATATATCAT	1260
AGAAAAACATA	TGTTTATATA	AAACAAATTT	TATCATATAT	GTGCAACTTA	ATACATGTGT	1320
ATATATATAT	AAACGCTAGC	AATGCACGGG	CATATAACTA			1380

GTATTAAGAT	GAATAAGAGG	GTATCCAAAT	AAAAAACTGG	TTGCTTACGT	ATGGATCGAA	1440
AGGGGTGGGA	AACGATTAAA	CGATTAAATC	TCTTCTTAGT	CAAAATTGAA	TAGAAGGAGA	1500
TTTAATATAT	CCCAATCCCC	TTGCATCATC	CAGGTGCAAC	CGTATAAGTC	CTAAAGTGCT	1560
GAGGAACACG	AAGAACCATT	GCATTGGCAT	GTAAGGCTCC	AGAAATTTGT	TGATCCTTAA	1620
ACAACATACA	GAACATCAAC	CAAAATTGCA	CGTCAAGGGT	ATTGGGTAAAG	AAACATCTAA	1680
ACAAATCTCT	TCGTGTGTGC	AAGRAACACG	GTGAGTCATG	CCGAGATCAT	ACTCATCTGA	1740
TATACATGCT	TACAGCTGCA	AAGACATTAC	AAACAACCTCA	TATTGCATTA	CAAGGATCTGT	1800
TTTATGAAAA	ATAAAATAGG	CCGGACACGA	CAAAAATCTCT	TGACGTGTAA	AGTAAATTTA	1860
ACAACAAAAA	AAAGCCATAT	GTCAAGCTAA	ATCTAATTCG	TTTTACGTAG	ATCAACAACT	1920
TGTAGAAGGC	AACAAACTAG	AGCCACGCAG	AAGTACAGAA	TGATTCCAGA	TGAACCATCG	1980
ACGTGCTACG	TAAAGAGAGT	GACGAGTCAT	ATACATTTGG	CAAGAAACCA	AGAGCTGTCC	2040
TACAGCCGTA	TCGCTGGCAT	AAGAACAACAA	GAATTTGGT	TAATTAATCA	TAAAGTATAA	2100
TAACGCTCGC	ATGCGCTGTC	ACTTCTCCAT	CACCACCACT	GGGTCTTCAG	ACCATTAGCT	2160
TTATCTACTC	CAGAGCGCGAG	AAGAACCCTGA	TCGACACCAT	GAAGTCGGTG	GAGAAGAAAC	2220
CGAAGGGTGT	GAAACAGAGT	CGCGGTGACA	AGCATAAGCT	GAAGACAGAG	TGGCCGGAGT	2280
TGGTGGGGAA	ATCGGTGGAG	AAAGCCAAGA	AGGTGATCCT	GAAGGACAAG	CCAGAGGCCG	2340
AAATCATATG	TCCTACGGTT	GGTACAAAAG	TGGGTAAGCA	TTATAAGATC	GACAAGGCTCA	2400
AGCTTTTGTG	GGATAAAAAG	GACAACTATC	CGCAGGTCCC	CAGGGTCGGC	TAGCCTCGAG	2460
ATCCCCGGCG	GTGTCCCCCA	CTGAAGAAAC	TATGTGCTGT	AGTATAGCCG	TCACCATGCA	2520
AGCTAGTTGA	GTCAATTTAG	GGCGATGATT	GTGATGAACA	ATTGAAGATG	AAAGAAAAAA	2580
TGGGTGGCAG	TTCTAGTTGT	AGCAATGACC	TTTAACCTTT	TATACAATTA	TGATATATG	2640
GTATTGTGCC	AAATTAACAG	TGTTATCATC	CATTTAGATA	TAGACGAAAA	AAAATCTAAG	2700
TTTTCTGTAT	ATGTCTAATT	TGAAATGAAG	GGAGTATATA	TGGGATATAA	TCGTGATAGA	2760
AACTAAAAAC	AATGCTAATT	TGAAATGAAG	GGAGTATATA	GTATCAGTGA	TACGTGTATT	2820
TCCCTCGTAA	TACTACCGAC	ATCACACGTC	TCCAGTTAAT	ACTATCAGAA	AGTCAACGGA	2880
CACATTTGTT	CGCGGTAGGC	GTACCCAAAC	ATTTTGATCG	GTTCCTTTTA	GTGAGGGTTA	2940
AGCGAGTCGA	CTTCGAGGGG	GGGCGCGGTA	CCCAGCTTTT	GTTCCTTTTA	TTATCCGCTG	3000
ATTGCGCGCT	TTCGCGTAATC	ATGGTCAATG	CTGTTTCTTG	TGTGAAATTT	TGCTTAATGA	3060
ACAATCTCAC	ACAACATACG	AGCCGGAAGC	ATAAAGTGTA	AAAGCTGGGG	GGCAAACTCG	3120
GTAGCTTAAC	TCACATTAAT	TGCGTTGCGC	TCACCTGGCG	CTTTTCCAGTC	GGGAACCTCG	3180
TCGTGCCACG	TGCATTAATG	AATCGGCCAA	GAGCGGGTTT	GAGCGGGTTT	CGGTATTGGG	3240
CGCTCTTCGC	TTCTCTCGCT	CACCTGACTC	CTGCGCTCGG	TCGTTCCGCT	GGCGGACAGG	3300
GTATCAGCTC	ACTCAAAGGC	GGTAATACGG	TTATCCACAG	AATCAGGGGA	TAAACGACGA	3360
AAGAACAATG	GAGCAAAAAG	CCAGCAAAAG	GCCAGGAACC	GTAAAAGGCG	CGCGTTGCTG	3420
GCCTTTTTC	ATAGGCTCGC	CCCCCTGAC	GAGCATCACA	AAAATCGAGC	CTCAAGTCAG	3480
AGGTGGCGAA	ACCCGACAGC	ACTATAAAGA	TACCAAGGCT	TTCCCTCTGG	AAAGCTCCCT	3540
GTGCGCTCTC	CTGTTCCGAC	CCTGCCGCTT	ACGGGATACC	TGTCGCGCTT	TCTCCCTCTG	3600
GGAAAGCTGG	CGCTTTCTCA	TAGCTCACGC	TCTAGTTCGT	TCAGTTCCGT	GTAGGTCGTT	3660
CGCTCCAAAG	TGGCGCTGTG	GCACGAAACC	CCCGTTTACG	CCGACCGCTG	CGCCTTATCC	3720
GGTAACATAT	GTCTTGAGTC	CAACCCGGTA	ATATCGCAGT	TATCGCTACT	GGCAGATGCC	3780
TGCGCTTAAC	ACCGCTACAC	TAGAAGGACA	GTATTGGTGA	CTACAGAGTT	CTTGAGGTGG	3840
GTACTTCTCG	GAATAAGAGT	TGGTAGCTCT	TGATCCGGCA	TCTGCGCTCT	CGCTGGTAGC	3900
GGTGGTTTTT	TTGTTTGCAA	CGCAGCAGTT	ACGCGCAGAA	AAAAGGATAT	CTCAAGAGAT	3960
CCTTTGATCT	TTTCTACGGG	GTCTGACGCT	CAGTGGAAAC	AAAACTCAGC	TTAAGGGATT	4020
TGGTCACTGA	GATTATCAAA	AAGGATCTTC	ACCTAGATCC	TTTTAAATTA	AAATTTGAAT	4080
TTTAAATCAA	TCTAAAGTAT	ATATGAGTAA	ACTTGTCTGT	ACAGTTTACCA	ATGCTTAACT	4140
AGTGAGGCAC	CTATCTCAGC	GATCTGTCTA	TTTGGTTTCT	CCATAGTTGC	CTGACTCCCC	4200
TCGTGTGAGA	TAACTACGAT	ACGGGAGGGC	TTACCATCTG	CGCCCATGTC	TGCAATGATA	4260
CGCGGAGACC	CACGCTCACC	GGCTCCAGAT	TTATCAGCAA	TAAACACGCG	AGCCGGGAGG	4320
GCCGACGCGA	GAGGTGGTCC	TGCAACTTTA	TCCGCTTCCA	TCCAGCTTAT	TAATTTGTTG	4380
CGGGAAGCTA	GAGTAAGTAG	TTGCGCAGTT	AATAGTTTGC	GCAACGTTGT	TGCCATTGCT	4440
ACAGGCATCG	TGGTGTACAG	CTCGTCTGTT	GGTATGGCTT	CATTGAGCTC	CGGTTCCCAA	4500
CGATCAAGGC	GAGTTACATG	ATCCCCATG	TTGTGCAAAA	AAGCGGTTAG	CTCCTTCGGT	4560
CCTCCGATCG	TTGTCAAGA	TAAAGTTGGC	GCAGTGTAT	CATCTATGCT	TATGGCAGCA	4620
CTGCATAATT	CTCTTACTCT	CATGCCATCC	GTAAGATGCT	TTTCTGTGAT	TGCTGAGTAC	4680
TCACCAAGT	TTTCTTGAGA	ATAGTGTATG	CGCGGACCGA	GTTGCTCTTG	CCGCGGCTCA	4740
ATACGGGATA	ATACCGCGCC	ACATAGCAGA	ACTTTAAAGG	TGCTCATCAT	TGGAACAGT	4800
						4860

TCTTCGGGGC GAAAACTCTC AAGGATCTTA CCGCTGTTGA GATCCAGTTC GATGTAACCC 4920
 ACTCGTGCAC CCAACTGATC TTCAGCATCT TTTACTTTCA CCAGCGTTTC TGGGTGAGCA 4980
 AAAACAGGAA GGC AAAATGCG CGCAAAAAAG GGAATAAGGG CGACACGGAA ATGTTGAATA 5040
 CTCATACTCT TCCTTTTTC ATATTATTGA AGCATTTATC AGGGTTATTG TCTCATGAGC 5100
 GGATACATAT TTGAATGTAT TTAGAAAAAT AACAAATAG GGGTCCGCG CACATTTCCT 5160
 CGAAAAGTGC CAC 5173

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 54 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

AGTATAAGTA AACACACCAT CACACCCTTG AGGCCCTTGC TGGTGCCCAT GGTG 54

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 55 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

CCTCATATCC CTTAGTGCCT AAGTTCGACG TCGGCCCTC TAGTCGACGG ATCCA 55

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

AGCGGAAAAA GCCCGAAAGG CTTCCCAAA TTGGC 35

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 45 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

TGCGCAGGCG TCTGCAAGTG TAAGCTGACT AGTAGCGGAA AATGC

45

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

TACAACCTTT GCAAAGTCAA AGGCGCCAAG AAGCTTTGCG CAGGCGTCTG

50

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 50 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: Other

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GCAAGAGTTG CTGCAAGAGT ACCCTGGGAA GGAAGTGCTA CAACCTTTC

50

The invention is not limited to the exact details shown and described, for it should be understood that many variations and modifications may be made while remaining within the spirit and scope of the invention defined by the claims.

47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100